

# 偏倚风险评估系列:(九)如何应用偏倚风险评估的结果

张渊 杨智荣 孙凤 詹思延

L8S 4L8 加拿大麦克马斯特大学临床流行病与卫生统计学系(张渊); CB1 8RN 英国剑桥大学临床医学院初级医疗中心(杨智荣); 100191 北京大学公共卫生学院流行病学与卫生统计学系, 北京大学循证医学中心(孙凤、詹思延)

通信作者:詹思延, Email: siyan-zhan@bjmu.edu.cn

DOI: 10.3760/cma.j.issn.0254-6450.2018.12.021

**【摘要】** 本文为偏倚风险评估系列讲座的最后一节,主要讨论偏倚风险评估结果的应用。偏倚风险评估是证据质量评估中关键的一环,能为循证决策提供重要的参考依据。偏倚风险评估结果可以用于单个研究的偏倚风险诊断,其结果可影响系统综述中对于原始研究的纳入排除和数据的分析。此外,本文重点讨论偏倚风险评估的结果如何应用到推荐分级的评价、制订与评估(GRADE)的整体证据质量评价,包括证据质量评价中的原则以及注意事项。

**【关键词】** 偏倚风险; 评估工具; 流行病学研究; 系统综述; 证据质量; GRADE

**基金项目:** 国家自然科学基金(81473067, 71673003)

**Risk of bias assessment: (9) Application of the risk of bias assessment results** Zhang Yuan, Yang Zhirong, Sun Feng, Zhan Siyan

Department of Clinical Epidemiology and Biostatistics, McMaster University, L8S 4L8, Canada (Zhang Y); Primary Care Unit, Department of Public Health and Primary Care, School of Clinical Medicine, University of Cambridge, Cambridgeshire CB1 8RN, UK (Yang ZR); Department of Epidemiology and Biostatistics, School of Public Health, Center of Evidence-based Medicine and Clinical Research, Peking University, Beijing 100191, China (Sun F, Zhan SY)

Corresponding author: Zhan Siyan, Email: siyan-zhan@bjmu.edu.cn

**【Abstract】** In this last paper of the series about risk of bias assessment, we introduce the application of risk of bias assessment results. Risk of bias assessment is one of the key steps in the assessment of quality of evidence. The risk of bias assessment results could be the “diagnosis” of individual studies, which helps decision making related to the inclusion and exclusion of individual studies, as well as the data analysis in the systematic review process. This paper focuses on how to incorporate risk of bias assessment results in the GRADE assessment for quality of evidence, including the principles and the tips for the application.

**【Key words】** Risk of bias; Tool for assessment; Epidemiological research; Systematic review; Quality of evidence; GRADE

**Fund programs:** National Natural Science Foundation of China (81473067, 71673003)

循证医学强调依靠最佳证据作出决策,而回答证据是否最佳关键的步骤就是偏倚风险的评估<sup>[1]</sup>。本系列讲座第2~8讲介绍了应用于各种不同情境,回应不同评估需求的偏倚风险评估工具,依次是随机对照试验(RCT)(平行设计<sup>[2]</sup>、交叉设计<sup>[3]</sup>、整群随机<sup>[4]</sup>)、非随机对照研究<sup>[5]</sup>、诊断试验<sup>[6]</sup>、预后研究<sup>[7]</sup>和系统综述<sup>[8]</sup>。本讲作为这一系列的结尾,将着重讨论如何使用偏倚风险评估的结果。理想状态下,医学决策是在了解证据的偏倚风险、评价证据质量的

基础上作出的。医学决策利益相关方主要包括但不限于卫生政策决策者、医学指南制定者、系统综述作者、卫生技术评估报告的作者以及这些文献和报告的读者,如临床医生、药师、护理人员 and 患者,同时也包括原始研究的设计者。本讲将重点从系统综述作者的角度来介绍如何应用偏倚风险评估结果,因为他们对于偏倚风险的评估最直接相关,不仅需要评估纳入的单个研究,还需要考虑根据评估结果对数据进行分析、探讨,从而得出结论;本文所涉及的讨

论也可为其他相关人员提供参考。

偏倚风险既可以按照结局进行评估和总结,也可以按照研究或者是偏倚风险的不同领域进行评估和总结。但前者最有意义,理由是:①决策一般是对某项干预措施的利弊进行权衡(balance between benefits and harms)之后作出的,在这里利和弊都是我们关心的结局,因此,决策时更关心的不是某个研究在整体上存在高还是低的偏倚风险,而是偏倚风险对于每一个结局的效应估计值的影响<sup>[9]</sup>;②在同一个研究中,同一个偏倚风险领域出现的缺陷或者问题,对于不同的结局产生的影响可能是不同的<sup>[1]</sup>,如盲法对客观和主观结局指标的影响。

一、偏倚风险评估结果的应用

以结局为单位进行偏倚风险的评估,按评估水平从低到高包括 3 个层次:第一个层次是事先确定评估领域的数目,通过回答每个领域的一系列信号问题,对单个研究在各个领域上的偏倚风险逐一作出评估(图 1 情境 a);第二个层次是基于多个偏倚风险领域对单个研究进行评估总结(图 1 情境 b1),也可以是某个特定偏倚风险领域上多个研究的评估总结(图 1 情境 b2);第三个层次则是针对某个特定结局的证据体中所有研究的偏倚风险评估总结(图 1 情境 c)。例如,在本系列讲座第 2 讲所举的实例中<sup>[2]</sup>,对于“偏离既定干预的偏倚”领域,被判定为低风险,

属于情境 a;而对于同一个实例,我们对多个领域评估之后,将研究的整体偏倚风险判定为低偏倚风险,属于情境 b1<sup>[2]</sup>。Cochrane 系统综述的撰写过程通常要求作者按照偏倚风险领域来总结纳入的研究,这就属于情境 b2。例如一项关于免疫疗法治疗水果过敏的系统综述中,综述作者纳入了两个临床试验,在“结果选择性报告”领域,作者判定为低偏倚风险水平;而在“结局测量”领域,作者判定为高偏倚风险水平,因为这两个临床试验均未采用盲法对结局“脱敏”和“免疫耐受”进行测量<sup>[10]</sup>。本文在稍后的段落中将介绍关于情境 c 的案例。

1. 在系统综述研究和撰写过程中的应用:偏倚风险的评估是系统综述中必不可少的环节,包括对个体研究的整体偏倚风险评价(情境 b1)以及某个评估领域的整体偏倚风险评价(情境 b2)。这样的评价结果,既可以影响系统综述的分析过程(决定是否纳入排除、亚组分析、敏感性分析和 Meta 回归),也有助于对于某项研究的优势或者不足进行充分讨论。

(1)纳入排除:在系统综述中排除高偏倚风险的研究,主要是出于对证据质量的考虑。尽管并不是所有的系统综述作者都在研究方案中明确将会排除具有高偏倚风险的研究,但已有越来越多系统综述者考虑将偏倚风险的评估作为文献入选标准中的一个条目。系统综述的制作并不总是严格遵循纳入研

研究	偏倚风险领域					
	随机化过程中的偏倚	偏离既定干预的偏倚	结局测量的偏倚	结局数据缺失的偏倚	结果选择性报告的偏倚	总结
研究1	高风险(情境a)	低风险(情境a)	低风险(情境a)	低风险(情境a)	低风险(情境a)	高风险(情境b1)
研究2	低风险(情境a)	高风险(情境a)	低风险(情境a)	低风险(情境a)	低风险(情境a)	高风险(情境b1)
研究3	低风险(情境a)	低风险(情境a)	高风险(情境a)	低风险(情境a)	低风险(情境a)	高风险(情境b1)
研究4	低风险(情境a)	低风险(情境a)	低风险(情境a)	高风险(情境a)	低风险(情境a)	高风险(情境b1)
研究5	低风险(情境a)	低风险(情境a)	低风险(情境a)	低风险(情境a)	高风险(情境a)	高风险(情境b1)
总结	? (情境b2)	? (情境b2)	? (情境b2)	? (情境b2)	? (情境b2)	证据体的偏倚风险(情境c)

图 1 偏倚风险评估结果应用的不同层次

究、证据评价、数据分析这一先后顺序,而是一个迭代牵制的过程。偏倚风险评估结果也可能会影响纳入排除和数据分析,在发现高偏倚风险可能导致证据质量降低的情况下,综述作者可以纳入或者排除一些研究,以获得高质量的证据。在理想条件下,综述作者应该遵循事先制订的研究方案来进行系统综述,因此,我们也建议综述作者在制订研究方案时,应该在数据分析中充分考虑偏倚风险给分析结果可能带来的影响,并且在纳入排除标准和数据分析中明确应对方案<sup>[11]</sup>。

(2)敏感性分析、亚组分析和Meta回归:在系统综述的数据分析过程中,我们需要考虑偏倚风险对于估计值的影响,所选择的策略包括敏感性分析、亚组分析和Meta回归等定量分析方法<sup>[12]</sup>。①敏感性分析:指通过检查研究方法、分析模型、未测量的变量或者研究假设的变化对研究结果的影响程度,来确定研究结果稳健性的方法<sup>[13]</sup>。在系统综述的实际操作中,综述作者常常通过调整某一个变量(纳入或者排除某项或多项研究),来观察这个变量对于估计值的影响。敏感性分析是分析具有某种偏倚风险特征的研究对于结果影响的常用手段。前面提到,系统综述作者可以根据偏倚风险评估结果来决定是否纳入或者排除某些研究,而利用敏感性分析方法则可以评价这些研究对于结果的稳健性的影响。敏感性分析方法不止应用于高偏倚风险的研究,也适用于由于结果的报告不全面而被判为“偏倚风险评估结果不明确”的研究。②亚组分析:根据某种特征,将纳入研究分为两个或者多个亚组,来观察亚组之间的异同<sup>[14]</sup>。在理论上,排除具有高偏倚风险的研究,仅呈现低偏倚风险的研究,等同于仅呈现低偏倚风险的亚组分析结果。此外,更为保守的策略是将纳入研究根据偏倚风险评估结果分成不同亚组,比较亚组之间的效应估计值的差异,分析某一特定亚组对于合并效应值的贡献,该策略也在我们随后讨论的证据质量评价中考虑高偏倚风险研究对于证据体的贡献时得到应用。③Meta回归:将某个变量(在此特指偏倚风险水平)作为回归分析的自变量,将效应估计值作为因变量,观察该变量(偏倚风险水平)对效应估计值的影响。使用Meta回归这种策略,假设一项系统综述使用OR来表示关联强度,那么系统综述作者可以采用相对比值比(ratio of odds ratios)来表示偏倚风险对于效应估计值的影响;类似的方法也适用于RR。相对比值比=高偏倚风险或者不明确偏倚风险的研究的比值比/低偏倚风险

研究的比值比<sup>[15]</sup>。

这3类策略的基本理念比较接近,但在具体操作方法、复杂程度以及结果呈现方式上存在不同。偏倚风险评估结果可以帮助系统综述作者决定哪些研究具有高偏倚风险,帮助作者决定哪些研究应该纳入敏感性分析、亚组分析和Meta回归以及如何纳入这些分析;另外一方面敏感性分析、亚组分析和Meta回归的结果均可以定量分析偏倚风险对于估计值的影响。例如,在一项研究中中年时期身体质量指数(body mass index, BMI)与痴呆的关联的系统综述<sup>[16]</sup>中,综述作者采用相对危险度来表示中年时期BMI与痴呆的关联强度。而为了探索纳入研究的特征对关联强度估计值的影响,研究者采用了Meta回归的方法,纳入的自变量包括几项相关的偏倚风险领域,比如选择偏倚、失访偏倚、结局确认偏倚等。研究者发现,原始研究若存在较高风险的选择偏倚、结局确认偏倚和失访偏倚,那么更可能得出中年时期低体重与阿尔茨海默症的关联。而控制研究的设计特征之后,这种关联就不再具有统计学意义。因此综述作者最后认为没有足够证据证实中年时期的低体重与痴呆的风险存在关联<sup>[16]</sup>。

(3)结果讨论:此外,系统综述的讨论部分可根据偏倚风险评估结果展开讨论,指出当前相关研究的不足,在此基础上提出对未来同类研究在设计和实施方面的启示。同时,评估结果也可以帮助阅读医学文献的读者客观理性地评判某项研究(而非系统综述中的证据体)的结果,而不是盲目选择相信某一项研究的结论。

2. 在评价系统综述和指南的证据质量中的应用:偏倚风险评估,是证据评价中非常重要的一环。目前,针对系统综述和指南制定过程中的证据评价,国内外广泛采用GRADE系统进行评价。GRADE工作组将系统综述中的证据质量定义为:对系统综述中效应估计值是否真实的信赖程度,或者说对系统综述中效应估计值是否真实的把握度<sup>[17]</sup>。系统综述作者需要考虑系统综述的证据质量,也就是说不但需要了解干预措施的疗效和副作用的大小(合并效应值,即用来估测真实疗效),同时还要了解证据的质量或者我们对效应估计值的信赖程度如何。GRADE将证据等级分为高、中、低、极低等4个级别,而如下5个因素均可降低我们对证据的信赖程度:纳入原始研究的偏倚风险、原始研究之间的一致性、证据的间接性、系统综述结果的不精确性和发表偏倚<sup>[17]</sup>。

(1)GRADE 证据评价的原则:GRADE 对于证据质量的评价是以结局为评价单位的。需要注意的是, GRADE 系统评价的是某个特定结局的证据体 (body of evidence), 而不是单个研究 (当然有可能某个证据体中只纳入了一项研究)(图 1 情境 c)。在偏倚风险这个领域, GRADE 要求使用者根据对纳入原始研究的偏倚风险评估, 对于某个结局的证据体的偏倚风险给出是非常严重、严重还是不严重的判断, 并据此相应地降低两级、一级或者不降低证据质量评价<sup>[18]</sup>。原则上, 对于结局的效应估计值, 若所涉及的高偏倚风险的研究越多, 则越有可能降低证据质量。对此, 我们可以通过检查在 Meta 分析中的权重来考量。较高偏倚风险的研究所占的权重越大, 就越需要降低证据质量。

(2)不同偏倚风险水平的估计值的接近程度:然而, 在考虑权重贡献的同时, 我们也要考虑来源于高偏倚风险研究的估计值和低偏倚风险研究的估计值是否接近, 这就需要用到我们之前讨论到的亚组分析方法。如果不同偏倚风险水平的估计值差异较大, 而且高偏倚风险的估计对于结局的总体估计值的贡献大于低偏倚风险的估计, 我们有理由认为总体估计值有可能受到高偏倚风险的影响, 并据此降低证据质量。一个特例是如果高偏倚和低偏倚风险的估计并无太大区别, 或者高偏倚风险的估计对于总体估计值的贡献较小, 此时 Meta 分析的合并估计值与低偏倚风险的估计并无太大区别。这时候我们可以不降低证据质量的等级。在这种情况下, 尽管有关这个结局的一部分数据受到高偏倚风险影响, 但是高偏倚风险并未导致有偏倚的估计 (这也是为什么我们要使用“偏倚风险”而不是“偏倚”的原因), 或者并未对证据体水平上的合并估计值产生重要的影响。例如, 在一项比较结构型甘油三酯与中、长链甘油三酯物理混合物用于外科或危重成年人患者肠胃外营养的系统综述中<sup>[19]</sup>, 综述作者比较了这两种不同的甘油三酯对于患者的蛋白质代谢和氮平衡,

以及血液生化指标。综述作者采用了 Cochrane 组织推荐的 RCT 偏倚风险评估工具评价纳入研究。在 10 项纳入研究 (共 526 名患者) 中, 有 4 项研究 (共 210 名患者) 被评定为低偏倚风险。在 10 项研究全部纳入的情况下, 两组 ALT 的合并均数差为 -4.39 IU/L (95%CI: -5.97 ~ -2.82), 提示使用结构型甘油三酯组患者 ALT 要比使用混合物患者低 4.39 IU/L。而仅纳入 4 项低偏倚风险研究的亚组分析, 提示均数差为 -3.95 (95%CI: -6.18 ~ -1.73)。纳入全部研究的分析结果, 与仅纳入低偏倚风险的研究非常接近, 尽管后者由于排除了一些研究, 置信区间变宽 (损失了部分精确性)。此时, 我们可认为证据存在“严重的偏倚风险”, 从而降低证据质量; 也可以选择不降低效应估计值的证据质量。后者实际上更合理, 这是因为尽管这些研究具有高偏倚风险, 而且对于合并效应值的权重贡献较大, 但是高偏倚风险的估计值和低偏倚风险的估计值非常接近, 可以推测偏倚风险并没有对效应值的估计产生较大影响。

(3)证据评价和偏倚风险评估结果的连续性:需要注意的是, 在证据质量评价中, 我们实际评价的是对证据的信心, 这个信心类似于一个“连续型变量”, 但是 GRADE 系统要求我们做出的判断是一个“分类变量”(高、中、低和极低); 同样是被评估为中等证据质量水平的证据, 对于有的中等证据, 我们虽然归为“中等”但是实际上我们可能认为这个证据质量已经几乎接近于低级证据了 (图 2 中的情境 a); 而对于另外一些中等证据, 我们对这些“中等证据”的信赖程度实际上要接近于高级质量的证据 (图 2 中的情境 b)。对于研究的偏倚风险的评价也类似, 有的存在偏倚风险 (some concern) 的研究, 实际研究质量整体水平可能接近一个虽然有些许瑕疵但是仍然被评为低偏倚风险的研究。不管是 GRADE 系统要求对证据体的偏倚风险做出的非常严重、严重和不严重的总体判断, 还是利用某个特定的偏倚风险工具, 例如 RoB2.0 版本<sup>[2]</sup>, 要求做出关于某个特定的研究的

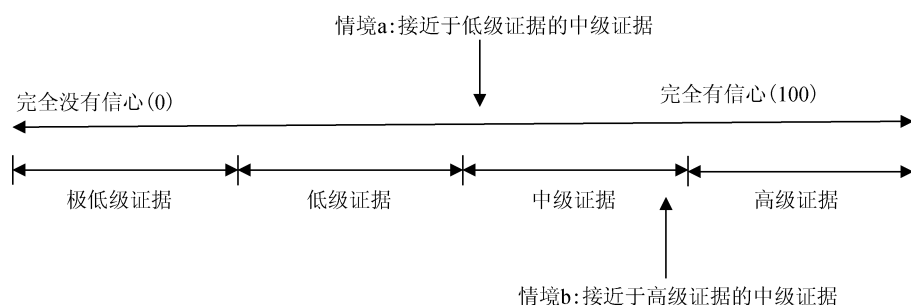


图2 证据评价是连续的

高偏倚风险(high risk of bias)、低偏倚风险(low risk of bias)或者存在偏倚风险(some concern)的判读,都是武断地、非精确地设置的分类变量,而不是真正反映现实世界中偏倚风险对结果影响程度由低到高的一个连续现象。

(4)不同的偏倚风险评估结果总结策略:在总结针对某个结局的偏倚风险时,可以选择情境b1的总结方式(图1),也可以选择情境b2的总结方式(图1),在特殊情况下不同的选择策略可能会造成最终的评价并不相同,例如图1中在相同的条件下,情境b1和情境b2两种不同的总结策略得出的结果是不同的,而依据这两种结果作出的GRADE评价也有可能不同。本文作者推荐选择情境b2的模式,这是因为证据质量评价更关注于某个偏倚风险领域(例如“结局的测量的偏倚”)上的问题是否对估计值产生了影响,而非笼统的证据体中纳入了多少高偏倚或者低偏倚风险的研究。在此,我们仍然强调系统综述的作者不应机械地应用偏倚风险评估工具和GRADE系统给出的标准,而是应该根据上述原则考虑高偏倚风险的研究对于证据体的贡献的大小,来考虑偏倚风险对估计值的影响,并据此做出是降低两级、降低一级或者是不降低证据质量的判断。通过以上关于高偏倚风险对于估计值的影响、证据评价的连续性、以及不同策略可能带来的影响,我们也可以看出,应用GRADE系统进行评价,不同的使用者并不能总是得出相同的评估结果。GRADE系统的设计者也已经注意到了这一点,他们认为GRADE系统不能总是帮助用户得出一致的判断,但是GRADE系统能够帮助用户系统化、结构化的评价证据,找出分歧的原因和依据,并且讨论解决分歧<sup>[20]</sup>。

(5)非随机干预性研究与GRADE系统潜在的“双重惩罚”:在使用非随机干预性研究的ROBINS-I<sup>[5]</sup>时,可能会存在一个“双重惩罚”的问题。这是因为传统的GRADE评价建议,对于有关干预效果的研究,证据评价等级的起点根据研究设计不同而不同:对于来源于RCT的证据,评价的起点是“高级证据”,而对于来源于非随机对照试验的证据,评价的起点是“低级证据”。这样的建议是基于流行病学方法学理论作出的,我们认为随机分配过程能够平衡基线时所有已知和未知的混杂因素,从而产生两个或者多个均衡可比的比较组。但是非随机干预性研究的偏倚风险评估工具中,这个因素对应于“混杂偏倚”和“研究对象选择的偏倚”这两个

领域,也就是说如果在非随机干预性研究的偏倚风险评估中,如果因为未能进行“随机分配”已将研究判定为“高偏倚风险”或者“极高偏倚风险”,那么再以“低级证据”作为证据质量评估的起点(已默认因为未能“随机分配”而降低了证据等级)就显得不合理,有“双重惩罚”的可能。所以,新的偏倚风险评估工具已经给GRADE系统的应用带来了挑战,新的GRADE指导意见发布指出,如果使用ROBINS-I评估非随机干预性研究,并且已经根据没有进行“随机分配”将研究判为“高偏倚风险”或者“极高偏倚风险”,之后再使用GRADE进行证据质量评价时,应该以“高级证据”作为起点,然后再根据ROBINS-I评估的结果,以及我们讨论到的GRADE评估的原则进行证据体的偏倚风险的判断,选择合适的降级,以避免在“低级证据”基础上的降级导致的对同一问题(没有随机分配)进行的“双重惩罚”<sup>[21]</sup>。

3. 促进研究质量的提升:除了应用于单个研究的评价和证据体的评价之外,偏倚风险评估工具的应用还可以在在一定程度上促进研究质量的提升。一个例子是在早期的随机对照试验研究中,很少有研究者注意到分配隐藏这一个细节。然而,无法做到分配隐藏的随机对照试验,尽管随机序列的产生没有问题,在“随机”完成后,干预实施过程中,针对医疗服务的提供者和研究对象都实施了盲法,仍然可能因为在“随机”过程中,由于有的研究者知道了“随机序列”,知晓了下一位研究对象可能分配在哪一个研究组中,因此仍然可能有研究者人为操纵分配的可能,从而导致随机的失败<sup>[5]</sup>。在偏倚风险评估工具提出这一个重要的考虑之后,开展随机对照试验的研究者们逐渐注意到了这个问题,因此,恰当地进行了分配隐藏的随机对照试验的比例逐渐升高。

## 二、讨论

1. 应用偏倚风险评估工具的目的:在偏倚风险或者研究质量的评价工具正式出现之前,流行病学家已经注意到了偏倚的可能及其对结果估计值可能的影响。传统流行病学教材中,一般将主要偏倚分为3类:选择偏倚、信息偏倚和混杂偏倚,而如果具体、翔实地列举所有可能的偏倚类型,则是一项极为繁冗复杂的工作,目前已知的偏倚种类超过数10种<sup>[22]</sup>。我们讨论偏倚风险的来源和分类,不是为了穷尽所有的研究过程中的缺陷和失误,而是为了确定可能给研究结果带来系统偏差的来源(偏倚风险的领域),并且评价这种风险性对于结果估计值的影响。所以,各式各样的偏倚风险或者研究质量评

价工具的目的旨在用一个更为结构化的模式来帮助使用者实现这一目的。重要的不是一个研究总体质量的高低,而是在不同领域上的偏倚风险方向及其对估计值的影响。既往评分式的评价工具如Jadad量表的缺陷就在于忽略了这一个重要的考量,而是试图量化各个领域并且通过总体得分给出一个总体量化评分<sup>[1]</sup>。这也提示我们应避免机械化地应用偏倚风险工具,不能期望通过偏倚风险评估工具精确评价研究质量,或者涵盖所有相关的偏倚风险来源。

2. 目前偏倚风险评估工具的缺陷:目前,我们在这个系列中所介绍的很多偏倚风险评估工具仍处于开发、更新和试用阶段。这些工具可能仍然存在信号问题设置不合理(重复、表述模糊、缺乏提示)、关键偏倚风险领域缺失、形成评价结果的标准不清楚等有待解决的问题。使用者在利用一项新开发的偏倚风险评估工具时,如果有疑问和反馈,可以联系开发者。

3. 与GRADE评价过程的重复:此外,部分偏倚风险评估工具可能包括了与GRADE系统重复的内容。除了之前我们提到的,如果使用ROBINS-I评价非随机干预性研究,且遵循GRADE建议将证据质量评价起点定为“低级”可能存在“双重惩罚”之外,更常见的一类重复是有的偏倚风险评估工具实际上已经可能包括了GRADE系统中需要考虑的间接性。例如,在诊断试验偏倚评估的QUADAS-2工具中的“适用性”<sup>[23]</sup>,系统综述偏倚风险评估的ROBIS工具中的“相关性”等<sup>[24]</sup>。这里特别需要厘清的是,原始研究所选择的样本及其目标人群,与系统综述作者和指南制定者所感兴趣的目标人群是不同的。原始研究所选择的样本对其目标人群的代表性决定着研究的内部真实性,而原始研究所选择的样本对于系统综述作者和指南制定者所感兴趣的目标人群的代表性,则决定着研究的外部真实性——前者属于内部真实性,属于偏倚风险的范畴;而后者外部真实性则属于GRADE评价过程中需要考虑的间接性,尽管很多研究者笼统地将这两个方面都称为“代表性(representativeness)”。因此,在将偏倚风险评估结果整合于GRADE系统进行证据评价时,尤其需要注意是否存在对于同一个问题的“双重惩罚”。例如在应用GRADE评价诊断试验的证据时,需要注意QUADAS-2工具中第一部分的“偏倚风险”与GRADE证据评价中的偏倚风险相关,而第二部分中的“适用性”与GRADE评价中的间接性相关<sup>[23]</sup>。而不应该根据QUADAS-2做出的两部分的评估结

果全部作为GRADE证据评价中偏倚风险的判断依据,并在此基础上再额外做出有关间接性的判断。

### 三、结论与展望

综上所述,应用偏倚风险评估工具的目的在于方便地、结构化地明确偏倚风险可能的来源,并且据此探索这个偏倚风险对于估计值的影响,而非涵盖研究中可能出现的问题和不足。有关偏倚风险的考量最终应纳入证据质量评价中,用于评价我们对于证据的信心,帮助我们更加明智、审慎地解读和应用证据,而不是盲目地根据已有证据进行决策。

利益冲突 无

### 参 考 文 献

- [1] 杨智荣,孙凤,詹思延. 偏倚风险评估系列:(一)概述[J]. 中华流行病学杂志, 2017, 38(7): 983-987. DOI: 10.3760/cma.j.issn.0254-6450.2017.07.027.  
Yang ZR, Sun F, Zhan SY. Risk of bias assessment: (1) overview [J]. Chin J Epidemiol, 2017, 38(7): 983-987. DOI: 10.3760/cma.j.issn.0254-6450.2017.07.027.
- [2] 杨智荣,孙凤,詹思延. 偏倚风险评估系列:(二)平行设计随机对照试验偏倚评估工具2.0介绍[J]. 中华流行病学杂志, 2017, 38(9): 1285-1291. DOI: 10.3760/cma.j.issn.0254-6450.2017.09.028.  
Yang ZR, Sun F, Zhan SY. Risk on bias assessment: (2) Revised Cochrane risk of bias tool for individually randomized, parallel group trials (RoB2.0) [J]. Chin J Epidemiol, 2017, 38(9): 1285-1291. DOI: 10.3760/cma.j.issn.0254-6450.2017.09.028.
- [3] 武珊珊,孙凤,詹思延. 偏倚风险评估系列:(三)交叉设计随机对照试验[J]. 中华流行病学杂志, 2017, 38(10): 1436-1440. DOI: 10.3760/cma.j.issn.0254-6450.2017.10.028.  
Wu SS, Sun F, Zhan SY. Bias risk assessment: (3) Revised Cochrane bias risk assessment tool for individual randomized, cross-over trials (RoB2.0) [J]. Chin J Epidemiol, 2017, 38(10): 1436-1440. DOI: 10.3760/cma.j.issn.0254-6450.2017.10.028.
- [4] 吕晓珍,孙凤,詹思延. 偏倚风险评估系列:(四)整群设计随机对照试验[J]. 中华流行病学杂志, 2018, 39(2): 240-244. DOI: 10.3760/cma.j.issn.0254-6450.2018.02.020.  
Lyu XZ, Sun F, Zhan SY. Risk related to bias assessment: (4) Revised Cochrane Risk of Bias Tool for cluster-randomized control trials (RoB2.0) [J]. Chin J Epidemiol, 2018, 39(2): 240-244. DOI: 10.3760/cma.j.issn.0254-6450.2018.02.020.
- [5] 孙凤,高乐,杨智荣,等. 偏倚风险评估系列:(五)非随机干预性研究[J]. 中华流行病学杂志, 2018, 39(3): 374-381. DOI: 10.3760/cma.j.issn.0254-6450.2018.03.024.  
Sun F, Gao L, Yang ZR, et al. Series of risk of bias assessment (5): Risk Of Bias In Non-randomized Studies-of Interventions (ROBINS-I) [J]. Chin J Epidemiol, 2018, 39(3): 374-381. DOI: 10.3760/cma.j.issn.0254-6450.2018.03.024.
- [6] 曲艳吉,杨智荣,孙凤,等. 偏倚风险评估系列:(六)诊断试验[J]. 中华流行病学杂志, 2018, 39(4): 524-531. DOI: 10.3760/

- cma.j.issn.0254-6450.2018.04.028.
- Qu YJ, Yang ZR, Sun F, et al. Risk on bias assessment: (6) A Revised Tool for the Quality Assessment on Diagnostic Accuracy Studies (QUADAS-2) [J]. *Chin J Epidemiol*, 2018, 39 (4) : 524-531. DOI: 10.3760/cma.j.issn.0254-6450.2018.04.028.
- [7] 唐少文, 张渊, 陶必林, 等. 偏倚风险评估系列: (七) 预后因素研究[J]. *中华流行病学杂志*, 2018, 39(7) : 1003-1008. DOI: 10.3760/cma.j.issn.0254-6450.2018.07.026.
- Tang SW, Zhang Y, Tao BL, et al. Risk of bias assessment: (7) Assessing Bias in Studies of Prognostic Factors [J]. *Chin J Epidemiol*, 2018, 39 (7) : 1003-1008. DOI: 10.3760/cma.j.issn.0254-6450.2018.07.026.
- [8] 胡晶, 孙凤, 詹思延. 偏倚风险评估系列: (八) 系统综述[J]. *中华流行病学杂志*, 2018, 39(8) : 1125-1129. DOI: 10.3760/cma.j.issn.0254-6450.2018.08.022.
- Hu J, Sun F, Zhan SY. Risk of bias assessment: (8) Risk of Bias in Systematic Review (ROBIS) [J]. *Chin J Epidemiol*, 2018, 39 (8) : 1125-1129. DOI: 10.3760/cma.j.issn.0254-6450.2018.08.022.
- [9] Guyatt GH, Oxman AD, Kunz R, et al. GRADE guidelines: 2. Framing the question and deciding on important outcomes [J]. *J Clin Epidemiol*, 2011, 64 (4) : 395-400. DOI: 10.1016/j.jclinepi.2010.09.012.
- [10] Yepes-Nuñez JJ, Zhang Y, Figuls RM, et al. Immunotherapy (oral and sublingual) for food allergy to fruits [J]. *Cochrane Database Syst Rev*, 2015, 11: CD010522. DOI: 10.1002/14651858.CD010522.
- [11] Shamseer L, Moher D, Clarke M, et al. Preferred reporting items for systematic review and Meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation [J]. *BMJ*, 2015, 349: g7647. DOI: 10.1136/bmj.g7647.
- [12] Katikireddi SV, Egan M, Petticrew M. How do systematic reviews incorporate risk of bias assessments into the synthesis of evidence? A methodological study [J]. *J Epidemiol Community Health*, 2014, 69(2) : 189-195. DOI: 10.1136/jech-2014-204711.
- [13] Schneeweiss S. Sensitivity analysis and external adjustment for unmeasured confounders in epidemiologic database studies of therapeutics [J]. *Pharmacoepidemiol Drug Saf*, 2006, 15 (5) : 291-303. DOI: 10.1002/pds.1200.
- [14] Sedgwick P. Meta-analyses: heterogeneity and subgroup analysis [J]. *BMJ*, 2013, 346: f4040. DOI: 10.1136/bmj.f4040.
- [15] Higgins JPT, Green S. *Cochrane handbook for systematic reviews of interventions* [M]. Chichester, West Sussex, UK: The Cochrane Collaboration, 2011.
- [16] Albanese E, Launer LJ, Egger M, et al. Body mass index in midlife and dementia: systematic review and Meta-regression analysis of 589 649 men and women followed in longitudinal studies [J]. *Alzheimers Dement (Amst)*, 2017, 8: 165-178. DOI: 10.1016/j.dadm.2017.05.007.
- [17] Guyatt GH, Oxman AD, Vist GE, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations [J]. *BMJ*, 2008, 336 (7650) : 924-926. DOI: 10.1136/bmj.39489.470347.AD.
- [18] Guyatt GH, Oxman AD, Vist G, et al. GRADE guidelines: 4. Rating the quality of evidence-study limitations (risk of bias) [J]. *J Clin Epidemiol*, 2011, 64 (4) : 407-415. DOI: 10.1016/j.jclinepi.2010.07.017.
- [19] Wu GH, Zaniolo O, Schuster H, et al. Structured triglycerides versus physical mixtures of medium-and long-chain triglycerides for parenteral nutrition in surgical or critically ill adult patients: systematic review and Meta-analysis [J]. *Clin Nutr*, 2017, 36(1) : 150-161. DOI: 10.1016/j.clnu.2016.01.004.
- [20] Guyatt G, Oxman AD, Akl EA, et al. GRADE guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables [J]. *J Clin Epidemiol*, 2011, 64 (4) : 383-394. DOI: 10.1016/j.jclinepi.2010.04.026.
- [21] Schünemann HJ, Cuello C, Akl EA, et al. GRADE Guidelines: 18. How ROBINS-I and other tools to assess risk of bias in nonrandomized studies should be used to rate the certainty of a body of evidence [J]. *J Clin Epidemiol*, 2018. DOI: 10.1016/j.jclinepi.2018.01.012.
- [22] Delgado-Rodríguez M, Llorca J. Bias [J]. *J Epidemiol Community Health*, 2004, 58 (8) : 635-641. DOI: 10.1136/jech.2003.008466.
- [23] Whiting PF, Rutjes AWS, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies [J]. *Ann Intern Med*, 2011, 155 (8) : 529-536. DOI: 10.7326/0003-4819-155-8-201110180-00009.
- [24] Whiting P, Savović J, Higgins JP, et al. ROBIS: a new tool to assess risk of bias in systematic reviews was developed [J]. *J Clin Epidemiol*, 2016, 69: 225-234. DOI: 10.1016/j.jclinepi.2015.06.005.

(收稿日期: 2018-01-30)

(本文编辑: 王岚)